© 2018 IJSRSET | Volume 4 | Issue 4 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
Themed Section : Engineering and Technology

Facial Expression Recognition using Convolutional Neural Networks

Chirag Ravat[1], Shital Solanki[2]
[1]M.E(I.T) Student, I.T Department, L.D College Of Engineering, Ahmedabad, Gujarat, India
[2]Assistant Prof., I.T Department, L.D College Of Engineering, Ahmedabad, Gujarat, India

ABSTRACT

Facial expression recognition by computer plays a key role in human computer interaction. FER has many applications such as Human-Robot interaction, surveillance, Driving-safety, Health-care, intelligent tutorial system, music for mood, etc. Basically, Facial expression recognition can be done using Artificial Neural Network (ANN) and Support Vector Machine (SVM). So the accuracy of facial expression depends on these two phases, Feature extraction phase and classification phase. So I've done survey on different methods of FER and even face recognition methods, and proposed a method to improve the accuracy of FER using Convolutional Neural Nets.

Keywords: Facial Expression Recognition, FER, Neural Networks, Convolutional Neural Networks.

I. INTRODUCTION

Artificial Neural Networks are computing systems inspired by the biological neural networks that constitute animal brains. Such systems learn to do tasks by considering examples, generally without task-specific programming. The human brain is composed of 86 billion nerve cells called neurons[1]. They are connected to other thousand cells by Axons. ANNs are composed of multiple nodes, which imitate biological neurons of human brain. The neurons are connected by links and they interact with each other.

The nodes can take input data and perform simple operations on the data. The result of these operations is passed to other neurons. The output at each node is called its activation or node value. Each link is associated with weight. ANNs are capable of learning, which takes place by altering weight values[1].

However Convolutional Neural Networks(CNNs) have an edge over conventional MLPs in terms of image recognition and classification. The brief introduction of CNN is given in chapter III.

II. FACIAL EXPRESSION RECOGNITION SYSTEM

Almost all prediction based systems consist of mainly two phases: Training phase and Testing phase. In either Supervised or Un-supervised learning method first feature vectors are trained based on given labelled data in case of Supervised learning or on given attributes in case of Un-supervised learning. These trained vectors are then used to test an unseen data and gives label accordingly.

In FER it is identified as Feature Extraction and Classification phases. In feature extraction phase certain features of an image is extracted based on which classifier, in classification phase, classifies image to one of the label of domain.

For fast and better training, first of all image is pre-processed to reduce noise. Pre-processing also

includes converting image into grayscale and resizing it into predefined dimensions. For different FER methods different sizes can be considered.

Then next step is to detect face from the pre-processed image. There are lots of techniques available for detecting face from an image like Haar classifier, Ada boost technique by Viola-jones, Adaptive skin colour, etc., which gives output as face image or non-face image[4].

After face is detected another step is to extract features from face image. It can be done directly from given image or also from video frames. Features such as eyes, nose, mouth, eyebrows, ears, etc. are detected. Feature extraction process works significantly well if face is already detected. There are two types of methods for feature extraction : Appearance based and Geometric based. Geometric based methods are more suitable for real time applications as Appearance based method consumes more power, time and memory but also highly discriminative. There are many feature extraction methods are there but Gabor feature and LBP[2] gives optimal results. There are also pre-defined Neural Net structures available such as AlexNet[6], which is a Convolutional Neural Net, useful in feature extraction.

In last classifier is used to train model using extracted features and classify them into label. Classifier learns the mapping between the given image and given label, so when we give a new image after sufficient training it can classify input image into certain label.
Also testing phase includes image pre-processing and feature extraction, based on these features classifier directly classifies that image into some label.

III. CONVOLUTIONAL NEURAL NETWORKS

Convolutional Neural Network (CNN) architecture is inspired by mammalian visual cortex. Visual cortex processes images in hierarchical manner, first low level features and then high level features. CNN also works same as visual cortex, it first processes low level features of an image, such as curves, edges, then bit higher features like small part of an image and this hierarchy is continued layer by layer and in last whole image is processed.

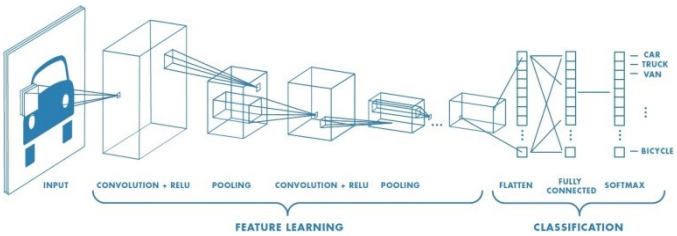


Figure 1. Basic CNN architecture

Basic layer of CNNs are convolutional layer, pooling layer, ReLU layer, Fully connected layer, loss layer, softmax layer, etc.

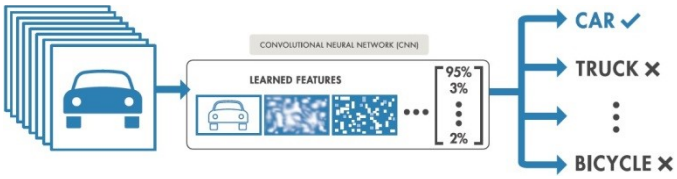


Figure 2. Result scheme of CNNs

Convolution layer takes an image as input and convolute it with feature vector or weight matrix and output more meaningful image. This image imported to other layer.

Pooling layer of CNN is used to extract most meaningful feature from each section of an image.
ReLU rectifies the image, it simply applied to check whether image gives some information or not. If image pixels are 0, the ReLU discards it.

Fully connected layer is used to sum up all learned features by connecting all neurons of previous layer to the next layer. It is usually used in later part of CNN architecture.

Loss layer in CNN, is used to apply different loss functions. For each attribute, a loss function can be applied. For example, softmax loss function is useful

in multi-class classification and gives output a label with probability.

IV. LITERATURE REVIEW of different FER methods that uses NN

FER using GRNN[2]

Method Used:- Viola jones for Face detection, LBP for feature extraction and GRNN for classification.

Advantages:- Accuracy Especially in Neutral, Happy, Surprise, classes (mostly 100%).

Disadvantages:- Confusion rates are higher in Unhappy, fear and angry classes. 62.5 % is the lowest achieved in angry class. Gives optimal solution only on 64×64 size block.

Facial Micro-expression recognition using Deep CNN[3]

Method Used:- Facial landmark detection using Dlib is carried out, CNN is used for face split in 12 ROIs, then HOOF features are calculated and SVM is used for classification.

Advantages:- Uses transfer learning, which reduces the training time. Uses more effective features, so accuracy is good.

Disadvantages:- Eye blinking in clip is ignored as noise. Efficient only for short duration clips.

FER using SVMs[4]

Method Used:- Haar cascades is for face detection, Fisher-face for distinguish faces, feature extraction for Dlib and SVM for classification is used.

Advantages:- Flow of method is good. Overall method takes significantly less time to classify image, Especially face detection part.

Disadvantages:- Noise removal is not efficient.

Smile detection using pair-wise distance vector and ELM[5]

Method Used:- CFAN for facial landmark detection, pair-wise distance vector for feature extraction and Extreme Learning Machine for classification is used.

Advantages:- ELM outperforms SVM, Adaboost, which gives better accuracy comparatively.

Disadvantages:- Running speed of algorithm is relatively slow, so real time implementation is difficult and cause more efficient hardware.

Gender and Age classification using CNNs[6]

Method Used:- Pre-trained CNN is used for feature extraction and SVM for classification is carried out.

Advantages:- Use of transfer learning through efficient pre-trained CNN – AlexNet gives very good results even dataset is small.

Disadvantages:- Even by using efficient feature extraction method, accuracy tents to low.

V. PROPOSED ALGORITHM

The proposed work is aimed to improve the accuracy of FER using supervised CNN splitting static image data in training and testing data.

Methodology

Step 1: Image pre-processing.

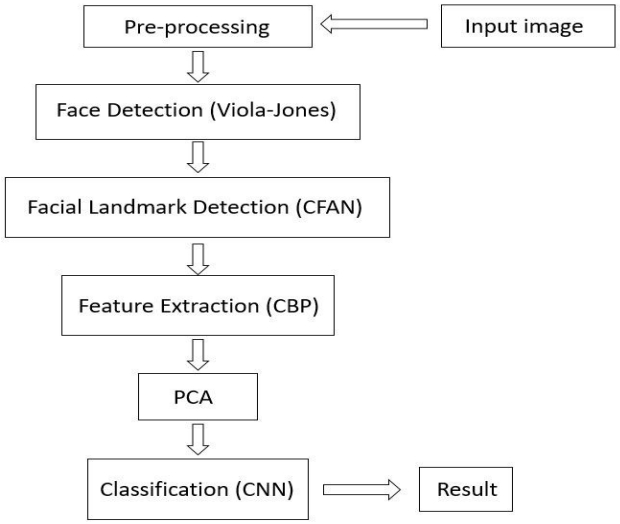
Step 2: Face detection.

Step 3: Facial Landmark Detection

Step 4: Facial feature extraction.

Step 5: Dimension Reduction.

Step 6: Classification using Convolutional Neural Network.



Flowchart For Proposed Model

Detailed Architecture of proposed CNN

Convolution Layer. 5×5 kernel, stride 1×1.
MaxPooling Layer of 3×3.
Convolution Layer. 5×5 kernel, stride 1×1.
MaxPooling Layer of 3×3.
Convolution Layer. 5×5 kernel, stride 1×1.
MaxPooling Layer of 3×3.
Convolution Layer. 2×2 kernel, stride 1×1.
Convolution Layer. 1×1 kernel, stride 11.
Fully Connected Layer.
Fully Connected Layer.
Softmax Function

First step is Image pre-processing, in which image is resized to 256×256 pixels and converted to gray scale. Then face detection is carried out using Haar-cascades to drop non face images, which contributes to lower false rates in final classification. After face detection is done, the detected face is cropped, so that only relevant portion of a picture is considered.

Facial landmark detection on that cropped face is done using Dlib ML toolkit. Using this facial landmarks, facial features are calculated using Central Binary Patterns.

In last classification is done using Convolutional Neural Network, which will consider features calculated from CBP and classify them accurately.

Here softmax regression function is used for result declaration.

VI. ANALYSIS OF PROPOSED METHOD

Implementation of proposed algorithm is carried out using Anaconda python. Instead of using small block sizes like 128×128 or 64×64, image starts with 256×256 block size, having more pixels to be inspected it gives relatively better results then smaller block sizes.

After detection of face, face is cropped in size of 156×156. Then it goes to whole process of feature extraction and classification.

Below is an example of face detection and cropping using Haar cascades and OpenCV.

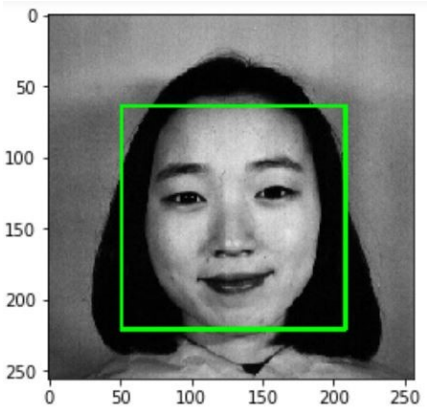


Figure 3. Face detection

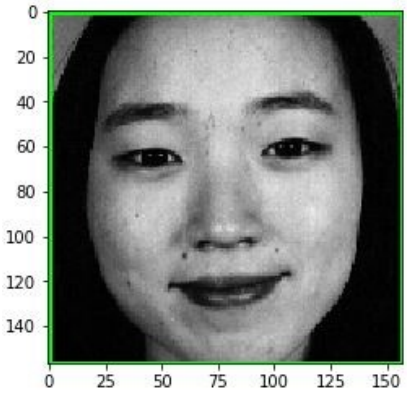


Figure 4. Cropping off Detected face

OpenCV, matplotlib are used for visualizing the processing on images.

Classification with our CNN model gives surprising accuracy of around 96%. But comes with a price. Since our CNN model is bit too large and feed by such high resolution images, its time complexity increases.

Comparison of our different testing:

Table 1

Method	Time	Accuracy
Simple CNN	Medium	62%
Simple CNN with Modified training scheme	Medium	76%
Our CNN model	High	88%
Our CNN with Modified training scheme	High	96%

In all of our model training, we got lowest of 81% accuracy, but after trained weight we got 96% in average accuracy.

We've trained our model on Jaffe and CK+ datasets. On Jaffe accuracy is lower then CK+. Training on Jaffe dataset gave 89-90% in average accuracy, while CK+ gives above 95%. It is because of number of images to be trained. Jaffe has only 213 static images, while CK+ has thousands of images, so model is trained well.

VII. CONCLUSION

Efforts have been done in the part of Feature extraction more than classification. Our proposed model takes high resolution data, so it takes bit more time to process an image but gives significantly higher accuracy than [2], [3] and [4], which have over 90% accuracy. It is trained and tested on both Jaffe and CK+ dataset, in both of them our model performed very well. Our future work will be focused on enhancing our algorithm and hence reducing the training time.

VIII. REFERENCES

[1]. Nikhil Buduma, 2015. Fundamentals of Deep Learning.2th Edn., Sebastopol, CA., ISBN: 978-1-491-92561-4

[2]. Kiran Talele, Archana Shirsat, Tejal Uplenchwar, Kushal Tuckley, "Facial Expression Recognition Using General Regression Neural Network", IEEE Bombay Section Symposium(IBSS), 2016.

[3]. Xiaohong Li, Jun Yu, Shu Zhan, "Spontaneous facial micro-expression detection based on deep learning" in IEEE, 2016.

[4]. Rajesh K M, Naveenkumar M, "A Robust Method for Face Recognition and Face Emotion Detection using Support Vector Machines", IEEE, 2016 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques(ICEECCOT), 2016.

[5]. Dongshun Cui, Guang-Bin Huang, Tianchi Liu, "Smile Detection Using Pair-wise Distance Vector and Extreme Learning Machine" in IEEE, 2016.

[6]. Xiaofeng Wang, Azliza Mohd Ali, Plamen Angelov, "Gender and Age Classification of Human Faces for Automatic Detection of Anomalous Human Behavior" in IEEE, 2017.

[7]. Jiaxing Li, Dexiang Zhang, Jingjing Zhang, Jun Zhang, "Facial Expression Recognition with Faster R-CNN", Science Direct, Procedia Computer Science 107 (2017) 135 – 140, 2017.

[8]. Nazima kauser, Jitendra Sharma, "Automatic Facial Expression Recognition: A Survey Based on Feature Extraction and Classification Techniques", IEEE, 2016.

[9]. Shubhada Deshmukh, Manasi Patwardhan, Anjali Mahajan, "Survey on real-time facial expression recognition techniques", IET (The Institute of Engineering and Technology) Journals, pp. 1-9, 2016.

[10]. Kewen Yan, Shaohui Huang, Yaoxian Song, Wei Liu, Neng Fan, "Face Recognition Based on Convolutional Neural Network", IEEE, Proceedings of the 36th Chinese Control Conference, 2017.